



**KALPRA ACADEMY**  
FOR SKILL DEVELOPMENT

# **JOB-ORIENTED 3-MONTHS ONLINE COURSE CURRICULUM**



# **DATA SCIENCE WITH PYTHON & AI**

+1 281-801-0921 [www.kalpracademy.com](http://www.kalpracademy.com) [Registration@kalpracademy.com](mailto:Registration@kalpracademy.com)

13111 Westheimer Rd., Suite 311, Houston, TX, 77077

## PROBABILITY FOR DATA SCIENTIST

- ▶ Basic Probability and Conditional Probability
- ▶ Properties of Random Variables
- ▶ Expectations (Mean) and Variance
- ▶ Entropy and cross-entropy
- ▶ Covariance and correlation
- ▶ Estimating probability of Random variable
- ▶ Understanding standard random processes.

## PROBABILITY DISTRIBUTIONS

- ▶ Normal Distribution
- ▶ Binomial Distribution
- ▶ Multinomial Distribution
- ▶ Bernoulli Distribution
- ▶ Probability, Prior probability, Posterior probability
- ▶ Bayes Theorem
- ▶ Naive Bayes
- ▶ Naive Bayes Algorithm
- ▶ Normal Distribution

## PYTHON PROGRAMMING

- ▶ Install python (Anaconda)
- ▶ Jupiter Notebook
- ▶ Install NumPy, Pandas,
- ▶ Matplotlib, Seaborn and SciKit Learn
- ▶ Spyder IDE

## **BASICS OF DATA PROCESSING**

- ▶ **Strings**
- ▶ **Lists**
- ▶ **Tuples**
- ▶ **Sets**
- ▶ **Dictionaries**
- ▶ **Control Flows**
- ▶ **Functions**
- ▶ **Formal, Positional, Keyword arguments**
- ▶ **Predefined functions (range, len, enumerates etc...)**

## **DATA FRAMES**

- ▶ **Packages required for data Science in R and Python.**
- ▶ **Packages required for data Science in R and Python.**

## **NUMPY PACKAGE**

- ▶ **One-dimensional Array**
- ▶ **Two-dimensional Array**
- ▶ **Pre-defined functions (arrange, reshape, zeros, ones, empty)**
- ▶ **Basic Matrix operations**
- ▶ **Scalar addition, subtraction, multiplication, division**
- ▶ **Matrix addition, subtraction, multiplication, division and transpose**
- ▶ **Slicing**
- ▶ **Indexing**
- ▶ **Looping**
- ▶ **Shape Manipulation**
- ▶ **Stacking**

## **PANDA'S PACKAGE**

- ▶ Series
- ▶ Data Frame
- ▶ Group By
- ▶ Crosstab
- ▶ Apply
- ▶ Map

## **DATA PREPARATION TECHNIQUES**

- ▶ Structured Data Preparation
- ▶ Data Type Conversion
- ▶ Category to Numeric
- ▶ Numeric to Category
- ▶ Data Normalization: 0-1, Z-Score
- ▶ Skew Data handling: Box-Cox Transformation
- ▶ Missing Data treatment

## **EXPLORATORY DATA ANALYSIS (EDA)**

- ▶ Statistical Data Analysis
- ▶ Exploring Individual Features
- ▶ Exploring Bi-Feature Relationships
- ▶ Exploring Multi-feature Relationships

## FEATURE ENGINEERING (FE)

- ▶ Combine Features
- ▶ Split Features

## DATA VISUALIZATION

- ▶ Bar Chart
- ▶ Histogram
- ▶ Box whisker plot
- ▶ Line plot
- ▶ Scatter Plot
- ▶ Heat Map
- ▶ Matplotlib, Seaborn– Visualization

## REGRESSION (SUPERVISED LEARNING)

- ▶ What is regression?
- ▶ Use Cases: Regression
- ▶ Linear Regression
- ▶ Theory behind Linear Regression
- ▶ Model Evaluation and related metrics
- ▶ Root Mean Square Error (RMSE)
- ▶ R-Square,
- ▶ Adj R-Square
- ▶ Feature selection methods
- ▶ Linear regression – Practice Problem

## OVERFITTING/UNDER FITTING

- ▶ Understand what overfitting is and under fitting model
- ▶ Visualize the overfitting and under fitting model
- ▶ How do you handle overfitting?

## DECISION TREES

- ▶ What are Decision Trees?
- ▶ Gini, Entropy criteria
- ▶ Decision trees in Classification
- ▶ Decision trees in Regression
- ▶ Ensembles
- ▶ Random Forest
- ▶ Boosting (Ada, Gradient, Extreme Gradient)
- ▶ SVM
- ▶ Ensembles

## RE-SAMPLING TECHNIQUES

- ▶ K-fold
- ▶ Repeated Hold-out Data
- ▶ Bootstrap aggregation sampling.

## TREE BASED ALGORITHMS

- ▶ Entropy
- ▶ Gini Index
- ▶ Information Gain
- ▶ Tree Pruning

## **CLASSIFICATION (SUPERVISED LEARNING)**

- ▶ What is Classification?
- ▶ Finding Patterns/Fixed Patterns
- ▶ Problems with Fixed Patterns
- ▶ Machine learning approach over fixed pattern
- ▶ Decision Tree based classification.
- ▶ Ensemble Based Classification
- ▶ Logistic Regression (SGD Classifier)
- ▶ Accuracy measurements
- ▶ Confusion Matrix
- ▶ ROC Curve
- ▶ AUC Score
- ▶ Multi-class Classification
- ▶ Softmax Regression Classifier
- ▶ Multi-label Classification
- ▶ Multi-output Classification.

## **ENSEMBLE MODELS**

- ▶ Random Forest
- ▶ Bagging
- ▶ Boosting
- ▶ Adaptive Boosting
- ▶ Gradient Boosting
- ▶ Extreme Gradient Boosting
- ▶ Heterogeneous Ensemble Models
- ▶ Stacking
- ▶ Voting

## **MULTIPLE/POLYNOMIAL REGRESSION (SCIKIT-LEARN)**

- ▶ Multiple Linear Regressions (SGD Regressor)
- ▶ Gradient Descent (Calculus way of solving linear equation)
- ▶ Feature Scaling (Min-Max vs Mean Normalization)
- ▶ Feature Transformation
- ▶ Polynomial Regression
- ▶ Matrix addition, subtraction, multiplication and transpose
- ▶ Optimization theory for data scientist.

## OPTIMIZATION THEORY (GRADIENT DESCENT ALGORITHM)

- ▶ Modelling ML problems with optimization requirements
- ▶ Solving unconstrained optimization problems
- ▶ Solving optimization problems with linear constraints
- ▶ Gradient descent ideas
- ▶ Gradient descent
- ▶ Batch gradient descent.
- ▶ Stochastic gradient descent.

## MODEL EVALUATION AND ERROR ANALYSIS

- ▶ Train/Validation/Test split
- ▶ K-Fold Cross Validation
- ▶ The Problem of Over-fitting (Bias-Variance trade-off)
- ▶ Learning Curve
- ▶ Regularization (Ridge, Lasso and Elastic-Net)
- ▶ Hyper Parameter Tuning (Grid Search CV)

## RECOMMENDATION PROBLEM

- ▶ What is Recommendation System?
- ▶ Top-N Recommender
- ▶ Rating Prediction
- ▶ Content based Recommenders.
- ▶ Limitations of Content based recommenders.
- ▶ Machine Learning Approaches for Recommenders.
- ▶ User-User KNN model, Item-Item KNN model
- ▶ Factorization or latent factor model
- ▶ Hybrid Recommenders
- ▶ Evaluation Metrics for Recommendation Algorithms
- ▶ Top-N Recommender : Accuracy, Error Rate
- ▶ Rating Prediction: RMSE.



## **CLUSTERING (UNSUPERVISED LEARNING)**

- ▶ Finding pattern and Fixed Pattern Approach
- ▶ Limitations of Fixed Pattern Approach
- ▶ Machine Learning Approaches for Clustering
- ▶ Iterative based K-Means Approaches
- ▶ Density based DB-SCAN Approach
- ▶ Evaluation Metrics for Clustering
- ▶ Cohesion, Coupling Metrics
- ▶ Correlation Metric.

## **SUPPORT VECTOR MACHINE (SVM)**

- ▶ SVM Classifier (Soft/Hard – Margin)
- ▶ Linear SVM
- ▶ Non-Linear SVM
- ▶ Kernel SVM
- ▶ SVM Regression.

## **PCA (UNSUPERVISED LEARNING)**

- ▶ Dimensionality Reduction
- ▶ Choosing Number of Dimensions or Principal Components
- ▶ Incremental PCA
- ▶ Kernel PCA
- ▶ When to apply PCA?
- ▶ Eigen vectors
- ▶ Eigen values.

## **MODEL DEPLOYMENT**

- ▶ Pickle (Pkl file)
- ▶ Model load from Pkl file and prediction.

## **ASSOCIATION RULES**

- ▶ A priori Algorithm
- ▶ Collaborative Filtering (User-Item based)

## **DEEP LEARNING ALGORITHMS DEEP LEARNING:**

- ▶ Introduction to Deep Learning
- ▶ Tensor flow
- ▶ Keras
- ▶ Setting up new environment for Deep Learning
- ▶ Perceptron model for classification and regression
- ▶ Perceptron Learning
- ▶ Limitations of Perceptron model
- ▶ Multi-layer Feed Forward NN model for classification and regression
- ▶ ML-FF-NN Learning with backpropagation.
- ▶ Applying ML-FF-NN and parameter tuning
- ▶ Pros and Cons of the Model
- ▶ Introduction to CNN with examples.

## **NATURAL LANGUAGE PROCESSING TEXT ANALYTICS (NLP):**

- ▶ Bag of words
- ▶ Glove Dictionary
- ▶ Text Data Preparation
- ▶ Normalizing Text
- ▶ Stop word and Whitespace Removal
- ▶ Stemming
- ▶ Building Document Term Matrix
- ▶ NLP (Natural Language Processing).

# MEET YOUR INSTRUCTORS



**Dr. Malleswar Yenugu**  
CEO of Kalpra Academy



**Suresh**  
Data Science Manager



**Partha**  
Senior Data Science Manager



**Sandesh**  
Data Science Manager